



20. konferenca Dnevi slovenske informatike

Sistem za avtomatsko gradnjo baze znanja iz spletnih mest



Ambrož Stropnik, KiviCOM d.o.o.

dr. Milan Zorman, UM-FERI

15. 04. 2013



slovensko
društvo
informatika



Vsebina

- Predstavitev podjetja
- Izhodišče
- Namen in cilji
- Platforma KiviKM
- Prikaz izdelane baze znanja
- SWOT analiza delovanja sistema
- Nadaljnji razvoj



Predstavitev podjetja



Sedež v Slovenj Gradcu, p.e. v Ljubljani

Ekipa

- 11 redno zaposlenih
- 5 stalnih zunanjih sodelavcev
- partner za sistemsko infrastrukturo in kolacije (Softnet)

**Prodaja večinoma v Sloveiji,
cca 15% izvoza**

Znanja

- Projektno vodenje
- Spletne tehnologije (predvsem MS. NET, html 5)
- Podatkovne baze (MS SQL, MySQL)
- Oblikovanje (CGP, razvoj blagovnih znamk, GUI aplikacije, mobile)
- Načrtovanje spletne prisotnosti, načrtovanje izvedbe aplikacij
- Oglaševanje in SEO
- Optimizacija poslovnih procesov



Predstavitev podjetja - reference





Izhodišče

- Spletni portali vsebujejo ogromno podatkov in informacij
- Podatki so v večini primerov zapisani v nestrukturirani obliki in berljivi le za ljudi
- Zahteve strank po vse bolj naprednemu iskanju informacij vse večje
- Primer naprednega iskanja v okviru portala podjetja Elektro energija d.o.o.:
 - „Kako prihraniti pri porabi električne energije?“
 - „Zakaj je moj račun električne energije tako visok?“

Rešitev: sistema za avtomatsko generiranje baze znanja iz spletnih mest



Namen in cilji

Namen

- Pridobiti neizkoriščene podatke (npr. blogi, novice, nasveti...) iz obstoječega portala (oz. portalov) in jih povzati v znanje
- Olajšati delo znanjskim delavcem pri izdelavi baz znanja
- Izboljšati iskalnik portala

Cilji

- Izdelava ustrezne metodologije oz. algoritma, ki bo z visoko stopnjo avtomatizacije izdelal ustrezno bazo znanja iz obstoječega portala
- Izdelana baza znanja mora biti izdelana na način, ki bo enostavno berljiv in nadgradljiv
- Možnost enostavne integracije izdelane baze znanja z drugimi (obstoječimi) bazami znanja



Platforma KiviKM

KiviKM – platforma za podporo procesom upravljanja z znanjem

Podpira celotni cikel upravljanja z znanjem, ki obsega:

- Pridobivanje znanja
- Kodiranje in shranjevanje znanja
- Prenos znanja
- Uporaba znanja

Razvita z Microsoftovimi tehnologijami (C#)

Jedro platforme predstavlja algoritem za avtomatsko granjo baze znanja iz obstoječega portala



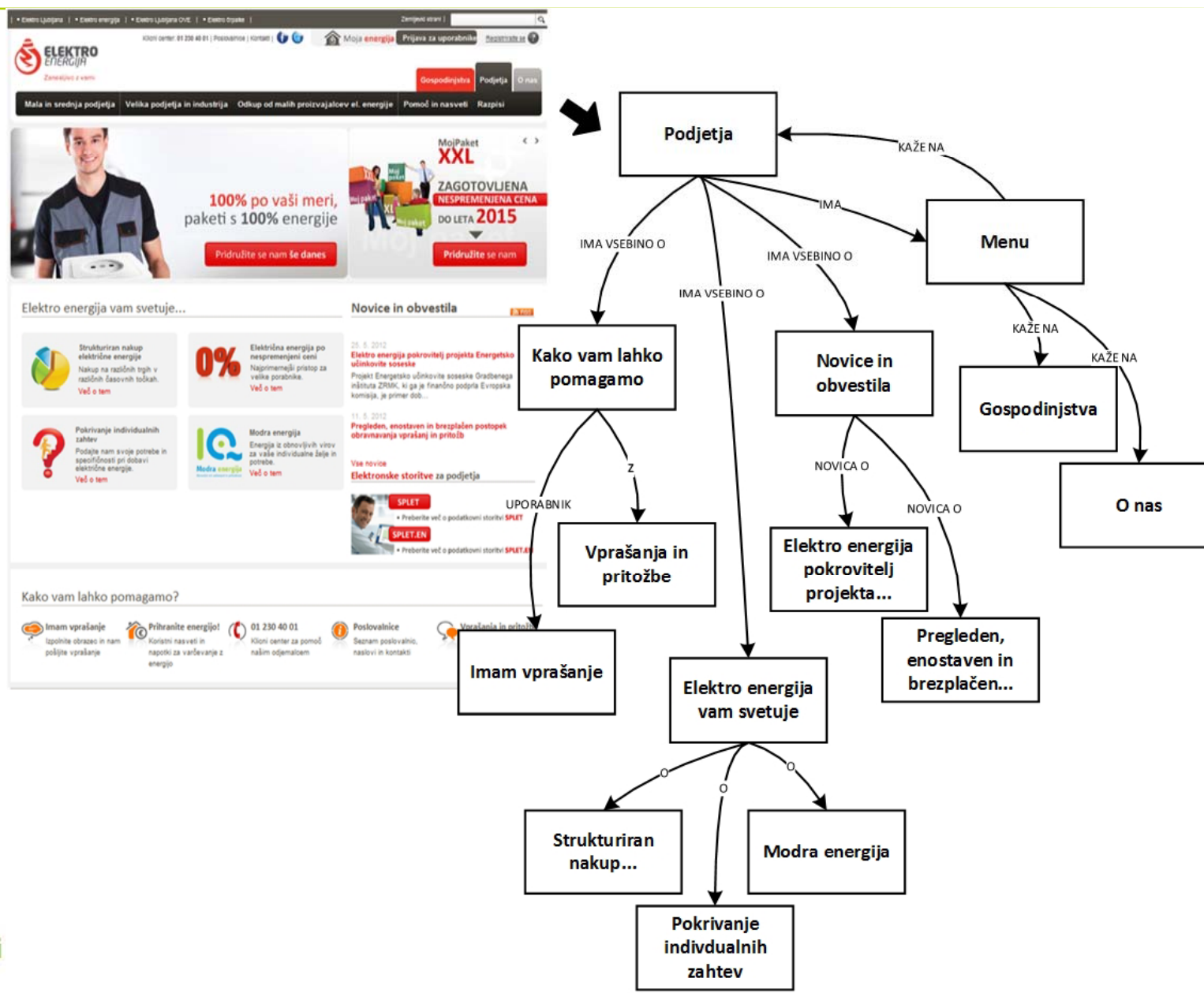
Algoritem za avtomatsko gradnjo baze znanja

Algoritem sestoji iz treh korakov:

- Vsako spletno stran razdelimo za segmente s pomočjo algoritma VIPS (angl. Vision-Based Page Segmentation)
- Rezultat segmentacije oz. dobljene segmente klasificiramo (Naive Bayes) v vnaprej določene razrede (npr. novica) in jih ustrezno semantično povežemo
- Povezane in klasificirane segmente predstavimo s pomočjo tehnologij semantičnega spleta - ontologije



Prikaz izdelane baze znanja





SWOT analiza delovanja sistema

Prednosti

- baza znanja izdelana na zelo podoben način kot bi jo izdelal znanjski delavec
- pridobitev podatkov in znanja iz različnih portalov
- za računalnik in človeka lažje berljivi podatki
- možnost široke uporabe

Priložnosti

- možnost povezave več baz znanj v eno skupno
- predlagane tehnologije omogočajo enostaven prenos znanja preko spleta

Slabosti

- na začetku velika odvisnost od znanjskega delavca
- počasna obdelava podatkov
- potreben nadzor s strani skrbnika oz. znanjskega delavca

Pasti

- uspešnost metode odvisna od znanjskega delavca
- možnost napak pri klasifikaciji segmentov
- verzije jezika HTML



Nadaljnji razvoj

Izboljšave algoritma in celotnega sistema

- Možnost izdelave baze znanja iz več portalov
- Razvoj metodologije za povečanje stopnje avtomatizacije pri klasifikaciji segmetov
- Pridobivanje znanja iz dokumentov (npr. pdf, doc...)
- Povezava sistema KiviKM z orodjem Protege (izdelava ustreznega vtičnika)

Postopno vključevanje sistema v uporabo pri novih in obstoječih projektih

- Nadgradnja lastnih iskalnikov z funkcionalnostjo iskanja po bazi znanja portala
- Nadgradnja self-service portalov s funkcionalnostjo svetovanja na osnovi baze znanja (npr. „Elektro energija“ oz. storitve v okviru „Moja energija“)



Hvala za vašo pozornost !

Vprašanja?

Pripombe?

Predlogi?